classified posts data store **330**. Finally, update corpus process **335**, updates the corpus used by the question answering system. Process **335** utilizes the classified posts found in classified posts data store **330** in order to update the corpus which is stored in corpus data store **340**.

[0039] Question answering pipeline **350** commences with question analysis process **355** that performs a question analysis using the updated corpus that is stored in corpus data store **340**. The question analysis results in one or more questions that most closely match the requested question. The closest matching questions are stored in closest matching questions data store **360**. Next, generate candidate answers process **365** generates candidate answers based on the questions that were identified by process **355**. The candidate answers are stored in candidate answers data store **370**. Rank/score candidate answers process **375** is performed to rank and score the candidate answers using a variety of factors, including supporting evidence that is found in the harvested discussions. The result of process **375** is a set of scored candidate answers that are stored in scored answers data store **380**. Finally, select answer process **385** selects the most likely correct answer based on the scores (e.g., the candidate answer with the highest score, etc.). The candidate answer that is the most likely correct answer is stored in selected answer data store **390** and is conveyed to a user of the question answering system as the most likely correct answer to the question posed by the user.

[0040] FIG. **4** is a depiction of a flowchart showing the logic used in site discovery of threaded online discussions and harvesting content from such discussions. The discussion site discovery process is shown commencing at **400** whereupon, at step **410**, the process continually crawls through the web using existing techniques to identify web sites **300** with threaded online discussion that can be harvested for use by the system. The threaded online discussion web sites, when found, are added to discussion sites data store **420**. As shown, step **410** is a continuous process that explores network **102**, such as the Internet, to identify such websites.

[0041] The discussion harvesting process is shown commencing at **430** whereupon, at step **440**, the process selects the first web site from discussion sites data store **420**. At step **450**, any new discussions not previously harvested are harvested (e.g., copied, collected, gathered, etc.). The harvested discussions are stored in harvested discussions data store **320** for further analysis and examination. A determination is made as to whether there are additional web sites that have been found by the site discovery process from which discussions need to be harvested (decision **460**). If there are additional web sites to process, then decision **460** branches to the "yes" branch which loops back to select the next web site from discussion sites data store **420** and harvest the new discussions which are added to harvested discussions data store **320**. This looping continues until all of the web sites from discussion sites data store **420** have been processed, at which point decision **460** branches to the "no" branch. At step **470**, the discussion harvesting process starts over with the first web site from discussion sites data store **420**. In this manner, discussions from new web sites found by the site discovery process are eventually harvested. In addition, new posts added to discussion threads are routinely captured and harvested after such new posts are added to their respective discussion threads.

[0042] FIG. **5** is a depiction of a flowchart showing the logic used to classify discussion posts and update a corpus utilized by a deep question answering system. Processing commences at **500** whereupon, at step **510**, the process selects the first harvested post from harvested discussions data store **320**.

[0043] At step **520**, the process employs a sentiment analysis process that analyzes the wording of the selected post. The analysis includes an analysis of the wording of the post, the grammar of the post, the structure of the post, and characters found in the post in order to identify the selected post as a question or an answer. A determination is made, based on the performed sentiment analysis, as to whether the selected post is a question (decision **530**). If the selected post is a question, then decision **530** branches to the "yes" branch whereupon, at step **540**, the text found in the post and the network identifier of the post (e.g., the uniform resource locator (URL), etc.) are added to corpus data store **340**. On the other hand, if the selected post is not a question, then decision **530** branches to the "no" branch bypassing step **540**.

[0044] A determination is made as to whether there are more posts in the set of harvested discussions to process (decision **550**). If there are more posts to process, then decision **550** branches to the "yes" branch which loops back to select and analyze the next post as described above. This looping continues until all of the posts in harvested discussions data store **320** have been processed, at which point decision **550** branches to the "no" branch. At step **560**, the process waits for new posts to be harvested and added to harvested discussions data store **320**. When new posts are added to harvested discussions data store **320**, the process loops back to select the newly added posts and process them as discussed above.

[0045] FIG. **6** is a depiction of a flowchart showing the logic performed by the question answering pipeline. Processing commences at **600** whereupon, at step **610**, the process receives a question from user **605** that the user desires the question answering system to provide a most likely answer. At step **620**, the process performs standard question analysis of the question received from the user. The standard question analysis includes a semantic analysis of the question. At step **630**, the process performs a primary search of corpus **340** in order to find questions previously identified in posts of online threaded discussions that are similar to the question that was received from the user. The questions from the corpus that most closely match the user's question are stored in closest matching question posts data store **360**. The questions in the corpus also include the network identifier (e.g., URL, etc.) that indicates the origin of the question from within the set of online discussions.

[0046] At predefined process **640**, the process generates a set of candidate answers that are stored in scored answers data store **380** (see FIG. **7** and corresponding text for further processing details). At step **680**, the process selects the candidate answer that has the highest score as the most likely answer to the user's question (selected answer **390**). The selected answer is returned to user **605** as being the most likely answer to the user's question.

[0047] FIG. **7** is a depiction of a flowchart showing the logic used by the system to generate candidate answers. Processing commences at **700** whereupon, at step **710**, the process selects the first question from closest matching question posts data store **360** with the selected question